Proceedings of the 51st Hawaii International Conference on System Sciences | 2018

Towards a Taxonomy of Data-driven Digital Services

Aya Rizk
Luleå University of Technology
aya.rizk@ltu.se

Birgitta Bergvall-Kåreborn
Luleå University of Technology
birgitta.bergvall-kareborn@ltu.se

Ahmed Elragal
Luleå University of Technology
ahmed.elragal@ltu.se

Abstract

Digitization is transforming every domain nowadays, leading to a growing body of knowledge on digital service innovation. Coupled with the generation and collection of big data, data-driven digital services are becoming of great importance to business, economy and society. This paper aims to classify the different types of data-driven digital services, as a first step to understand their characteristics and dynamics. A taxonomy is developed and the emerging characteristics include data acquisition mechanisms, data exploitation, insights utilization, and service interaction characteristics. The examined services fall into 15 distinct types and are further clustered into 3 classes of types: distributed analytics intermediaries, visual data-driven services, and analytics-embedded services. Such contribution enables service designers and providers to understand the key aspects in utilizing data and analytics in the design and delivery of their services. The taxonomy is set out to shape the direction and scope of scholarly discourse around digital service innovation research and practice.

1. Introduction

The reach and heterogeneity of the current information and communication technologies (ICT) is swiftly leading to the digitization of our everyday interactions. Digitization, that is the disruptive effects of digital technology in today's organizations and society [1], is giving way to the growing body of knowledge on Digital Service Innovation (DSI). Organizations are thus turning to digital services to increase their competitiveness, growth and innovativeness [2], [3]. The adoption and use of such services results in the generation of big data about people using it or surrounding objects [4]. This phenomenon has been coined in the literature as datafication [1]. Big data is typically characterized by the 3Vs - increasing volumes, velocity and variety – but can also be associated with its analytical techniques or the value it generates [5], [6].

Thus, the two phenomena of digitization and datafication are intertwined and reinforcing one another [1]. Thus, there is a need to study digital services that rely on big data and analytics [4], [7]. The extant literature on digital services, including their design and evaluation methods and descriptions, fails to incorporate this "language of data" [4, p. 17]. This paper is motivated by understanding data-driven digital services, enabling the design of new and improved services and advancing the DSI domain.

Hence, the aim of this paper is to understand the types and characteristics of data-driven digital services in relation to the utility of data, analytics and insights. One way to build and structure this knowledge is through systematic examination and classification of existing data-driven digital services into a taxonomy. While taxonomy development is well-recognized in the IS literature [8], the current discourse on digital service taxonomies lacks the perspective of big data and analytics, and in particular their utility in services. To this end, we pose the following research questions *"What characterizes data-driven digital services in relation to the utilization of data and analytics?"* and *"How can data-driven digital services be clustered?"*

The research questions are answered by developing a taxonomy in two iterations: the first goes from conceptual characteristics of services, verified by empirical cases, and the second examines empirical cases to refine the conceptual characteristics. We have focused our attention to digital services with a high degree of open general access, mobility, pervasiveness, sensing, and need for data as a value creating resource.

The proposed taxonomy aims to support service designers by highlighting how they can utilize data and analytics for service design and delivery. It also helps digital infrastructure providers understand the needs of service designers using their digital infrastructure, particularly those needs pertained to data and analytics. The remainder of the paper is organized as follows. Section 2 reviews related research on digital service and data-driven taxonomies. Section 3 describes the method, followed by section 4 that outlines the emerging characteristics

URI: http://hdl.handle.net/10125/50022
ISBN: 978-0-9981331-1-9

of data-driven digital services. Section 5 then presents the clustered types before discussing the results and concluding the paper in section 6.

2. Related Research

Classification methods have long been foundational to scientific inquiry. Bailey [9] differentiates between taxonomies and typologies arguing that taxonomies are developed empirically while typologies are conceptually grounded. However, the two terms are often used interchangeably in the literature. Especially in the IS field, we find that taxonomic classifications are sought due to the practical relevance and empirical evaluation they provide. A number of taxonomic classifications have been proposed in the digital and ICT-enabled service realm. However, none was found to classify data-driven digital services. The following list includes some of the most relevant classifications in the literature, with regards to the scope of this paper.

One of the first taxonomies relevant for digital services are Steven Alter's [10] taxonomy of decision support systems. While the concept of a digital service has dramatically changed during the past 40 years, the nature of decision support systems certainly leaves traces of relevance to taxonomies of data-driven services. More recently, Williams et al. [11] proposed a taxonomy for digital services focusing on their underlying design principle, and contributing to the understanding and assessment of design dimensions and service provider objectives in relation to digital services.

A fast growing subclass of digital services is mobile services and a few taxonomies related to mobile applications and m-commerce have emerged [12]–[14]. For instance, Nickerson et al. [8] develop a taxonomy for mobile applications driven by the meta-characteristic of high-level end-user interaction. The characteristics of mobile applications are accordingly represented by temporal, communication, transaction, access and location dimensions.

Going from mobile to ubiquitous computing and its datafication effect, further taxonomies emerged to accommodate the changing nature of digital services and their value creating business models. Hartmann et al.'s [15] taxonomy of data-driven business models highlights the role of big data and analytics in service value creation. Even though their taxonomy is built upon the business model literature, it is relevant to this paper because of the role of data as a key resource in value creation. Unlike the previously discussed taxonomies, Hartmann et al.'s [15] taxonomy did not set a meta-characteristic to drive their selection of characteristics. Instead, quantitative analysis allowed them to identify those characteristics with highest discriminatory power when clustering data-driven business models. Data source and key activity were then defined as the two main clustering variables, grouping the examined business models into 7 types. Their taxonomy targets service designers and providers, as well as business managers since it uses data as a key innovation resource and analytics to enhance the value creation.

There is also a need to expand the service perspective beyond service designers and providers and include the users of digital services. Taking the position of the service user while developing a taxonomy is crucial in understanding what types of services may be lacking and what opportunities arise [16]. Lee & Lee [16] propose a citizen-centric taxonomy with characteristics derived from marketing and service science domains. Built on citizen-centricity, the four characteristics forming this taxonomy are the mode of technology, service purpose, service authority and delivery mode. With nine values under these characteristics, 17 smart city service categories were formed.

Table 1. Summary of related taxonomy research

Taxonomy	Meta-characteristic	Target audience	Ref.
Mobile apps	User-application interaction	Researchers and mobile app developers	[8], [13]
mCommerce innovations	Co-creation empowerment	Service providers & partners	[14]
Pervasive mobile apps	Application requirements	Application and middleware developers	[12]
Digital services	Design & service provider objectives	Service designers	[11]
Smart city services	Citizen-centricity	Urban planners, administrators & service designers	[16]
Data-driven innovative start-ups	--	Start-up firms & service providers	[15]

A summary of the taxonomies described above, their meta-characteristics and target audience is presented in Table 1. Since existing taxonomies related to digital services do not address data-driven digital services from a data and analytics utility perspective, we build on the extant literature on digital service taxonomies in order to propose a taxonomy for data-driven digital services.

3. Research Method

Taxonomic classification refers both to the process of classification and the output classification model [9]. Nickerson et al. [8], [13] have developed a method for taxonomy development for IS to enable researchers classify artifacts. Despite their focus on digital artifacts we find the method general enough to be used for digital services, especially considering the tight coupling between many digital products and services. Hence, this paper follows this taxonomy development method in two iterations.

At the outset of the study, we defined the taxonomy's meta-characteristic – derived from the research gap presented earlier – as the utilization of data and analytics in digital service. Then, we specified ending conditions as advised by Nickerson et al. [8], where the taxonomy needs to be:

- Subjectively; concise, extendible, comprehensive and explanatory;
- No new dimensions or characteristics have emerged in the last iteration; and
- No objects were merged or split into one or more classes, respectively, in the last iteration.

In addition, we apply Bailey's [9] classification rule in which a taxonomy needs to be exhaustive, where each of the classified objects have been assigned to a suitable class. The first iteration followed a conceptual-to-empirical approach where the initial set of dimensions and characteristics were first conceptualized based, primarily, on characteristics manifested in the literature and then related to a sample of existing services. A purposeful sampling technique was applied to select a sample of 10 data-driven services [17], which were then examined against these characteristics and their possible values. The sampling technique focused on digital services with a high degree of mobility, pervasiveness, sensing, and need of data for added functionality. Secondary data about these services has been gathered from their official websites, blogs and news articles. The outcome from this iteration is the initial version of the taxonomy presented in the next section.

The second iteration followed an empirical-to-conceptual approach, where an additional set of data-driven services were identified from an ongoing co-creative smart city initiative [18]. Primary data collected through interviews with service leads was also used in this iteration. Data on each service was analyzed to code the different characteristics and the values each service hold for each characteristic. Common characteristics were used to revise the taxonomy. Furthermore, a cluster analysis was conducted on all services to examine a high-level pattern for the emerging types. The analysis was conducted using RapidMiner 7.3 using a nominal distance measure and K-means clustering. Since the four characteristics act as variables, it was necessary to have at least 16 (2^4) services to classify [19].

Appendix A provides an overview of the service types, and Appendix B provides service descriptions of selected services, due to space limitations. This revised taxonomy, along with the clusters, is presented in section 5 *Types of data-driven digital services*.

4. Characteristics of data-driven digital services

This taxonomy is anchored in the utilization and flow of data, information and knowledge in digital services. In the literature, the process of discovering knowledge from data is typically referred to as knowledge discovery in databases, or KDD [20]. However, with the volume, variety and velocity of data generation going beyond the capacity of databases, the term has evolved into data analytics [21], [22]. The taxonomy's four main characteristics (see Figure 1) follow the value chain of big data and extracted knowledge [20], [23], [24]. However, KDD and more recent big data analytics processes [25] end with interpreting the results, typically by a human decision maker. In data-driven digital services, generated insights were found to be utilized in different ways, which in turn manifested different modes of service interaction. Hence, in addition to data acquisition and data exploitation, two characteristics specific to digital services have emerged: insights utilization and service interaction. The following subsections detail each characteristic and its possible corresponding values, *italicized* in the text.

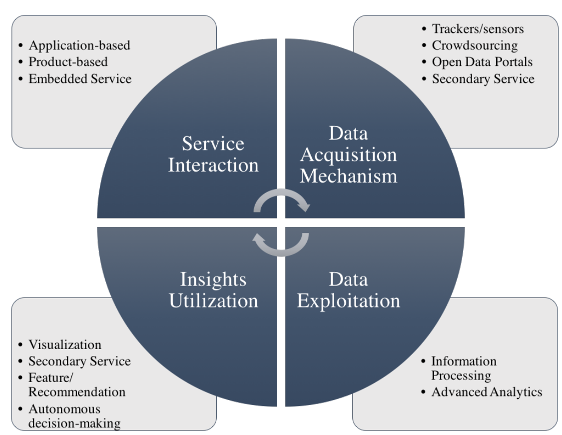

Figure 1. Characteristics of data-driven digital services

4.1. Data acquisition mechanism

Data-driven digital services are usually designed to make use of a) datasets generated through the service use, and/or b) pre-existing datasets. The former category can be divided into two subcategories: data generated through tracking/sensing, or crowdsourcing, differentiated based on the active participation of the user. Existing datasets can be divided into openly available data and data generated by other services. It is also becoming more common that services rely on data acquired via different modes, combining some of the following sources.

Trackers and/or Sensors is a dominant mechanism in services that rely mostly on tracking users or events, or measuring specific phenomena through sensing. Tracking can either be done through web tracking (e.g. using cookies) or via other devices (e.g. GPS in smartphones), which in most cases is acquired on a user level of granularity. Sensing, on the other hand, uses sensory technology to measure phenomena on the individual user, event, or regional level [15].

Crowdsourcing refers to the sourcing of a task to a set of contributors or a distributed group of people, instead of assigning it to a designated agent [26], [27]. In this context, the task involves data generation and/or annotation, which is acquired through the digital service. The difference between this mode of data acquisition and tracking/sensing is the active participation of the user sharing the data. Even though in both cases users should provide their consent to sharing, oftentimes in crowdsourcing data acquisition takes place at the foreground of the service, being part of or interfering with the service experience.

Open data portals refers to freely available data that can be accessed, reused and republished for different purposes without limitations or restrictions [28], [29]. It is usually structured, cleansed and available in formats that are machine readable and downloadable [30]. Even though governments and public organizations are considered the largest creators and publishers of open data, other entities such as scientific institutions, private organizations and individuals can publish it [29].

The last mechanism of data acquisition is one in which a service depends on data generated through *secondary services*, and not publicly or freely available for access. For example, a service like Data Waze [31] was found to depend on its sister service Waze for data acquisition, and build narrative models and visualizations accordingly. Similarly, Facebook and Google provide their targeted advertisement services based on data collected through their social media and search services, respectively.

4.2. Data exploitation

This section describes the different processing and analytical activities employed on the data acquired that adds value, as well as the types of outcomes from each activity. In the literature, there is a long tradition of presenting this collection of activities as a linear process; such as the process of knowledge discovery in databases [20], insights extraction from big data [32], or the big data value chain [23], [24]. Data represents the lowest abstraction level in the knowledge pyramid, when describing events or facts [21], [33]. The aim of exploitation is to extract information and knowledge that represent new insights and create additional value from this data [20], [21]. Exploitation of (big) data and analytics can take then two values: information processing, and advanced analytics.

Information processing represents a relatively lower degree of abstraction of data as it was generated or acquired. This encompasses various activities known in the literature as data transformation, cleaning and processing [15], aggregation [32], synthesizing [34], or transformation [20]. The outcome from this activity is still quite detailed, but well processed for insights utilization. Services that utilize processed data typically does so on the same level of the acquired data. For example, a service for crowdsourcing traffic incidents displays the processed data on the same level (i.e. street or road) as the reported contributions. Information processing in this case entails cleansing and verification. On the other hand, aggregation refers to the formation of a whole from the collection of constituent units [35]. In this context, units can either refer to similar variables for different instances or different variables for similar instances. Thus, there are two key types of aggregations observed in the examined services. The first type is aggregating values of the same variables towards a lower level of detail, and the outcome of this type of aggregation is typically computed on a different, whole, unit of analysis. The second type of aggregation is one where variables for similar instances are combined from different sources [15], also referred to as data integration [36].

Advanced analytics refers to the suite of methods, tools and techniques applied on (big) data to extract useful and hidden patterns and relationships, or more generally insights [22], [37], [38]. A key difference between advanced analytics and information processing is that the latter exploits insightful information while the former extracts previously unknown knowledge. Advanced analytical techniques are rooted in data mining, statistics, machine learning, visual analytics and other domains [39]. These techniques have been typically applied to structured data in database management systems (DBMS) and data warehouses, but with the rise of big data, the techniques evolved to cope with unstructured data, sensory and mobile content, and computational distribution [40]. The knowledge extracted through advanced analytics can be classified into descriptive, predictive, prescriptive knowledge [15].

4.3. Insights utilization

This characteristic deals with how the insights (information or knowledge) are utilized within the service. In KDD and big data analytics processes, this activity is referred to as visualization, distribution and access [15], dissemination [32], distribution [34] or interpretation [20]. Even though the most common understanding is to distribute or disseminate the generated insights, signaling that insights are perceived to be in the form of reports or documented knowledge, more recently there has been an increasing call for “operationalizing” insights, where insights should be embedded in workflows, processes and decisions [32]. Thus, we refer to this dimension as Insights Utilization, where different services have insights utilized differently.

Kambatla et al. [41] highlight three key enabling roles of analytics that apply to digital service configurations. First, insights guide effective decision-making through *visualization* [38]. Even though visualization of data and insights are very valuable for domain experts interpreting them, they are usually of little value to a wider user base. Thus, service providers, researchers, journalists and others have tended to storytelling, or narrative visualizations [42], to enrich visualizations for better communication of data and insights [43].

Second, insights can be operationalized to enact a service *feature* or a *recommendation*. In interactive or user-centered environments, the role of insights is to guide the user-service interface, such as with smart applications and recommendation-based services [41]. In this context, insights are operationalized to shape the service interaction, and this is manifested in our sample in the form of enacting a service feature (e.g. locating a PokéStop in Pokémon GO) or explicitly making a recommendation (e.g. suggesting a customized biking route in Calimoto).

Third, insights may guide the design and evolution of complex systems [41], primarily through *autonomous decision-making*. This form of insights utilization closes the loop of data and analytics flow in a service interaction, and allows for continuous learning and evolution of such service systems. Fourth, insights can interface with a *secondary service* – physical or digital –including providing data-as-a-service and analytics-as-a-service offerings [3], [44].

4.4. Service interaction

This characteristic describes the modes of service interaction, and what technologies the interaction entails. Williams et al. [11] discussed a similar notion that is service delivery requirements, where high requirements indicated the user’s need for specialized hardware or software, and low requirements indicated that older resources would work fine for the particular service delivery. Service users may interact with digital services in different ways.

A very common and unsurprising service interaction happens through a dedicated smartphone or smart watch application, designed specifically for such a service, which utilizes a device commonly owned by the potential user. Service designers may also lower the interaction barrier and utilize existing applications to host their service. For example, more than 34,000 chat bots are now active through Facebook Messenger [45], each acting as its own digital service [46]. Furthermore, some digital services’ interactions take place through web applications. In this case, a web browser is all that is needed to interact with the service. Collectively, this mode of interaction is referred to as *application-based interaction*.

It is also common to find emerging digital services bundled with specific physical products, usually called cyber-physical services or product-service systems [47], [48]. The interaction with such service typically takes place through the product, so we refer to this mode of interaction as *product-based interaction*. The degree of interaction with the product differs from one service to another. For example, self-driving cars require minimum interaction from the service user (passenger), since they continuously learn from contextual data. On the other hand, an interactive visualization toolbox (e.g. Public Like Displays) imposes higher levels of interaction with the user.

The interaction with a data-driven digital service may also take place through another, digital or physical, encapsulating service and its underlying infrastructure. For example, IBM’s Deep Thunder provides a data-driven digital service that forecasts weather on a hyper-local and short-term levels [49]. In turn, a local service provider puts the insights into action through the city infrastructure, such as in the case of Rio de Janeiro’s Operations Centre [50]. Similarly, New York’s Fire Control Department made use of a catastrophic risk model that predicted whether inbound city complaints were related to severely bad living conditions in the Bronx and Brooklyn – an indicator that correlated highly with fire hazards [51].

Whether the secondary service is a digital or a physical service, the focal data-driven service interaction is considered to be embedded within a wider service interaction. Thus, we refer to this mode as *embedded service* interaction. In the following section, we discuss the different types of data-driven services emergent through this taxonomy.

5. Types of data-driven digital services

This section describes different types of data-driven digital services based on the above-mentioned characteristics and selected from the second iteration. Given the identified characteristics and values, there are 96 possible types of data-driven services. Our sample – shown in the appendix - shows 15 different types of services that could be further grouped into three high-level clusters. Three clusters provided us with a reasonable distribution and cohesion given our relatively small sample. The first cluster, labeled *distributed analytics intermediaries,* includes 4 service types that acquire data primarily through crowdsourcing, perform advanced analytics, interface their insights with secondary services and that are most commonly interacted with as an embedded service. The insights generated by these types of services are used in a secondary physical or digital service, including providing DaaS and AaaS offerings within a different service exchange.

The second cluster covers 6 service types primarily characterized by their information processing activities, and utilization of insights through visualizations. In addition, all services that are driven by open data belong to this cluster. Visualization and storytelling play a key role in communicating generated insights to the user within a given service interaction. Thus, this cluster of services is interpreted and labeled as *visual data-driven services*. The third cluster contains 5 service types that are driven by advanced analytics but rely more on tracking and sensing data sources. In addition, most of these services utilize the generated insights within the service interaction, either through a feature, recommendation or autonomous decision-making. This cluster is labeled as *analytics-embedded services*.

In this given classification, the dimensions with highest information gain (IG) ratios are *data exploitation* (1.0) and *insights utilization* (0.6), which means they are more influential in classifying data-driven services than *data acquisition mechanisms* and *service interaction* mode (0.3 and 0.2, respectively). This suggests that the choice of analytical techniques and putting insights to action is crucial to data-driven digital service, which is consistent with the literature [32]. In addition, the fewer examples of product-based interaction indicate that there is an opportunity for more product-service offerings that are data and analytics-driven. Similarly, services that apply autonomous decision-making present an opportunity for innovative data-driven services where machine learning and artificial intelligence techniques are required and embedded.

6. Discussion

By identifying 4 characteristics and 13 values, we identify 96 (combinations) possible types of data-driven digital services. Our sample manifested 15 of them that could be further clustered into three patterns: distributed analytics intermediaries, visual data-driven services, and analytics-embedded services. In terms of data acquisition for digital services, tracking through service use and sensing are highly represented in our sample. One reason for this preferred data acquisition mechanisms is the degree of control in setting what data to collect during the service interaction. On the other hand, open data is comparably less trustworthy in terms of completeness and accuracy [52]. The current classification highlights the importance of (narrative) visualizations to service users. The proposed taxonomy also highlights opportunities for product-service offerings making use of advanced analytics and autonomous decision making, given the advancements in sensing and tracking technologies.

The purpose of this paper is to explore the different types of data-driven services, which represent the main unit of analysis. The proposed taxonomy and its four characteristics have been generated by examining service instances, from initiation following the data and analytics utilization to service interaction. While the traditional objective of data and analytics is supporting decision-making processes in organizations, the rise of digitization and big data is pushing its boundaries to include unfamiliar terrains. This is evident where more services are offered and used beyond organizational contexts with a wide and open user base [2], [41]. Existing taxonomies either take a service designer perspective [11], provider and business model perspective [15], or a user perspective [16]. To the best of our knowledge, this is the first taxonomy to classify data-driven digital service taking a dichotomous perspective in order to follow the utilization of data and analytics.

7. Conclusion and future research

The proposed taxonomy enables service designers and providers to understand the key characteristics in utilizing data and analytics in the design and delivery of their innovative services. The emerging service types show the rising opportunities and challenges of studying digital services using this language of data [4]. The increasing embeddedness of analytics in service design and the automated utilization of insights has its implications on both service and big data analytics research. On one hand, service researchers and designers need to tackle data as an increasingly valuable resource, while processing and analyzing it towards utilization that serves the experience with minimal complexity. One the other hand, big data vendors and researchers have the opportunity to develop tools and methods that help service designers elucidate data and analytics more effectively.

Service structures are not addressed in this paper. To follow a biological analogy, some services are compound, while others are quite atomic. In this paper, this issue is not addressed in detail, and for simplicity, a service is considered at the level of interaction with a user or another service. Moreover, the services examined to guide this taxonomy are data-driven digital services that are used in an urban context. Hence, the taxonomy could benefit from further iterations and testing using more data-driven services from other contexts. Moreover, understanding whether and how these different types impact the innovation of data-driven digital service at large is a future avenue for research in DSI.

Acknowledgment

This work was funded by the European Commission under grant no. 645198 OrganiCity and we are grateful for the contribution which made this research possible.

References

[1] C. Loebbecke and A. Picot, "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda," *J. Strateg. Inf. Syst.*, vol. 24, no. 3, pp. 149–157, Sep. 2015.

[2] H. Demirkan, C. Bess, J. Spohrer, A. Rayes, D. Allen, and Y. Moghaddam, "Innovations with smart service systems: analytics, big data, cognitive assistance, and the internet of everything," *Commun. Assoc. Inf. Syst.*, vol. 37, no. 1, p. 35, 2015.

[3] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud," *Decis. Support Syst.*, vol. 55, no. 1, pp. 412–421, Apr. 2013.

[4] P. P. Maglio and C.-H. Lim, "Innovation and Big Data in Smart Service Systems," *J. Innov. Manag.*, vol. 4, no. 1, pp. 11–21, 2016.

[5] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact.," *MIS Q.*, vol. 36, no. 4, pp. 1165–1188, 2012.

[6] A. McAfee and E. Brynjolfsson, "Big data: the management revolution," *Harv. Bus. Rev.*, vol. 90, no. 10, pp. 60–68, 2012.

[7] A. Rizk, B. Bergvall-Kåreborn, and A. Elragal, "Digital Service Innovation Enabled by Big Data Analytics-A Review and the Way Forward," in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[8] R. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in information systems," *Eur. J. Inf. Syst.*, vol. 22, no. 3, pp. 336–359, May 2013.

[9] K. D. Bailey, *Typologies and taxonomies: an introduction to classification techniques*, vol. 102. Sage, 1994.

[10] S. Alter, "A taxonomy of decision support systems," *Sloan Manage. Rev.*, vol. 19, no. 1, p. 39, 1977.

[11] K. Williams, S. Chatterjee, and M. Rossi, "Design of Emerging Digital Services: A Taxonomy," in *Design Research in Information Systems*, vol. 22, Boston, MA: Springer US, 2010, pp. 235–253.

[12] K. M. Dombroviak and R. Ramnath, "A taxonomy of mobile and pervasive applications," in *Proceedings of the 2007 ACM symposium on Applied computing*, 2007, pp. 1609–1615.

[13] R. Nickerson, J. Muntermann, U. Varshney, and H. Isaac, "TAXONOMY DEVELOPMENT IN INFORMATION SYSTEMS: DEVELOPING A TAXONOMY OF MOBILE APPLICATIONS," 2009.

[14] L. Khansa, C. W. Zobel, and G. Goicochea, "Creating a Taxonomy for Mobile Commerce Innovations Using Social Network and Cluster Analyses," *Int. J. Electron. Commer.*, vol. 16, no. 4, pp. 19–52, Jul. 2012.

[15] P. M. Hartmann, M. Zaki, N. Feldmann, and A. Neely, "Big data for big business? A taxonomy of data-driven business models used by start-up firms," 2014.

[16] J. Lee and H. Lee, "Developing and validating a citizen-centric typology for smart city services," *Gov. Inf. Q.*, vol. 31, pp. S93–S105, Jun. 2014.

[17] H. Suri, "Purposeful sampling in qualitative research synthesis," *Qual. Res. J.*, vol. 11, no. 2, pp. 63–75, 2011.

[18] OrganiCity, "Organicity," *OrganiCity*, 2015. [Online]. Available: http://www.organicity.eu/. [Accessed: 06-Apr-2017].

[19] E. Mooi and M. Sarstedt, *Cluster analysis*. Springer, 2010.

[20] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in knowledge discovery and data mining*, vol. 21. AAAI press Menlo Park, 1996.

[21] R. Kitchin, *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.

[22] P. Russom, "Big data analytics," *TDWI Best Pract. Rep. Fourth Quart.*, vol. 19, p. 40, 2011.

[23] H. G. Miller and P. Mork, "From data to decisions: a value chain for big data," *IT Prof.*, vol. 15, no. 1, pp. 57–59, 2013.

[24] E. Curry, "The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches," in *New Horizons for a Data-Driven Economy*, Springer, 2016, pp. 29–37.

[25] O. Müller, I. Junglas, J. vom Brocke, and S. Debortoli, “Utilizing big data analytics for information systems research: challenges, promises and guidelines,” *Eur. J. Inf. Syst.*, vol. 25, no. 4, pp. 289–302, 2016.
[26] Gartner, “IT Glossary - Crowdsourcing,” *Crowdsourcing*, 2017. [Online]. Available: http://www.gartner.com/it-glossary/crowdsourcing. [Accessed: 10-Apr-2017].
[27] J. Howe, *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House, 2008.
[28] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “Dbpedia: A nucleus for a web of open data,” *Semantic Web*, pp. 722–735, 2007.
[29] T. Jetzek, M. Avital, and N. Bjorn-Andersen, “Data-driven innovation through open government data,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 9, no. 2, pp. 100–120, 2014.
[30] E. Lakomaa and J. Kallberg, “Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs,” *IEEE Access*, vol. 1, pp. 558–563, 2013.
[31] “Data Waze,” *Data Waze*, 2014. [Online]. Available: https://data-waze.com/. [Accessed: 10-Apr-2017].
[32] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, “Big data, analytics and the path from insights to value,” *MIT Sloan Manag. Rev.*, vol. 52, no. 2, p. 21, 2011.
[33] D. Rosenberg, “Data before the fact,” in *“Raw Data” is an Oxymoron*, Cambridge, Mass.: MIT Press, 2013, pp. 15–40.
[34] J. F. Rayport and J. J. Sviokla, “Exploiting the virtual value chain,” *Harv. Bus. Rev.*, vol. 73, no. 6, p. 75, 1995.
[35] “Aggregate,” *Merriam Webster Online*. .
[36] M. Lenzerini, “Data integration: A theoretical perspective,” in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2002, pp. 233–246.
[37] R. Agarwal and V. Dhar, *Editorial—Big data, data science, and analytics: The opportunity and challenge for IS research*. INFORMS, 2014.
[38] N. Elgendy and A. Elragal, “Big data analytics: a literature review paper,” in *Industrial Conference on Data Mining*, 2014, pp. 214–227.
[39] A. Elragal and M. Haddara, “Big Data Analytics: A text mining-based literature analysis,” in *Norsk konferanse for organisasjoners bruk av IT*, 2014, vol. 22.
[40] H. Chen, R. H. Chiang, and V. C. Storey, “Business intelligence and analytics: From big data to big impact.,” *MIS Q.*, vol. 36, no. 4, pp. 1165–1188, 2012.
[41] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, “Trends in big data analytics,” *J. Parallel Distrib. Comput.*, vol. 74, no. 7, pp. 2561–2573, 2014.
[42] E. Segel and J. Heer, “Narrative visualization: Telling stories with data,” *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1139–1148, 2010.
[43] R. Kosara and J. Mackinlay, “Storytelling: The next step for visualization,” *Computer*, vol. 46, no. 5, pp. 44–50, 2013.
[44] R. Schüritz and G. Satzger, “Patterns of Data-Infused Business Model Innovation,” in *Business Informatics (CBI), 2016 IEEE 18th Conference on*, 2016, vol. 1, pp. 133–142.
[45] C. O’Brien, “Facebook Messenger chief says platform’s 34,000 chatbots are finally improving user experience,” *VentureBeat*, 2016. [Online]. Available: http://venturebeat.com/2016/11/11/facebook-messenger-chief-says-platforms-34000-chatbots-are-finally-improving-user-experience/. [Accessed: 20-Mar-2017].
[46] Colour-in City, “http://colourincity.com/,” 2017. [Online]. Available: http://colourincity.com/. [Accessed: 20-Mar-2017].
[47] M. Barrett, E. Davidson, J. Prabhu, and S. L. Vargo, “Service innovation in the digital age: key contributions and future directions,” *MIS Q.*, vol. 39, no. 1, pp. 135–154, 2015.
[48] A. Neely, “Exploring the financial consequences of the servitization of manufacturing,” *Oper. Manag. Res.*, vol. 1, no. 2, pp. 103–118, 2008.
[49] IBM, “IBM100 - Deep Thunder,” 2012. [Online]. Available: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/. [Accessed: 21-Mar-2017].
[50] A. M. Townsend, *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company, 2013.
[51] M. Flowers, “Beyond Open Data: The Data-Driven City,” *Transpar. Open Data Future Civ. Innov.*, pp. 185–198, 2013.
[52] S. Kim and J. Lee, “E-participation, transparency, and trust in local government,” *Public Adm. Rev.*, vol. 72, no. 6, pp. 819–828, 2012.
[53] MobiliCityX, “MobiliCity,” *MobiliCityX*, 2016. [Online]. Available: https://www.mobilicity.london. [Accessed: 03-Sep-2017].
[54] E. Field, “Colour-in City,” *“Hey there! Is this a good time for a chat?,”* 2017. [Online]. Available: http://colourincity.com/2017/01/25/hey-there-is-this-a-good-time-for-a-chat/. [Accessed: 02-Sep-2017].
[55] AirPublic, “Where the wind blows,” 2017. [Online]. Available: https://medium.com/@AirPublica/where-the-wind-blows-3bd8871d9c07. [Accessed: 02-Sep-2017].
[56] “Waze,” 2017. [Online]. Available: https://www.waze.com. [Accessed: 02-Sep-2017].
[57] “Public Like Displays,” *Public visualizations in Santander*, 2017. [Online]. Available: http://rxd.architectuur.kuleuven.be/organicity/index.php/2017/02/23/public-visualizations-at-santander/. [Accessed: 02-Sep-2017].
[58] “Calimoto,” *Calimoto*, 2017. [Online]. Available: https://calimoto.eu/en/. [Accessed: 02-Sep-2017].
[59] “Crackberry,” *Real-time traffic information with Google Maps*, 2007. [Online]. Available: https://crackberry.com/real-time-traffic-information-google-maps. [Accessed: 02-Sep-2017].
[60] “Google Official Blog,” *Stuck in traffic?*, 2007. [Online]. Available: https://googleblog.blogspot.se/2007/02/stuck-in-traffic.html. [Accessed: 09-Jun-2017].

Appendix A – Types of data-driven digital services

Cluster	Services	Data Acquisition Mechanism	Data Exploitation	Insights Utilization	Service Interaction
Cluster 1: Distributed analytics intermediaries	MobiliCity	Crowdsourcing	Advanced Analytics	Secondary Service	Application
	NY Housing Fire Control; Eldesmarque InstaSport; Colour-in City	Crowdsourcing	Advanced Analytics	Secondary Service	Embedded Service
	AirPublic; WearAQ	Trackers/Sensors	Advanced Analytics	Secondary Service	Embedded Service
	TranquilCity	Crowdsourcing	Information Processing	Secondary Service	Embedded Service
Cluster 2: Visual data-driven services	Waze	Crowdsourcing	Information Processing	Visualization	Application
	CityMapper; SpendNetwork	Open Data Portal	Information Processing	Visualization	Application
	InCityTogether	Trackers/Sensors	Information Processing	Visualization	Application
	Public Like Displays (PLD)	Open Data Portal	Information Processing	Visualization	Product
	SmartHalo	Trackers/Sensors	Information Processing	Visualization	Product
	Breathable Cities	Trackers/Sensors	Information Processing	Visualization	Embedded Service
Cluster 3: Analytics-embedded services	Data-Waze	Secondary Service	Advanced Analytics	Visualization	Application
	Pokémon GO; Calimoto	Trackers/Sensors	Advanced Analytics	Feature/Recommendation	Application
	Google Traffic	Trackers/Sensors	Advanced Analytics	Visualization	Embedded Service
	FitStar	Trackers/Sensors	Advanced Analytics	Feature/Recommendation	Product
	Waymo	Trackers/Sensors	Advanced Analytics	Autonomous Decision Making	Product

Appendix B – Service descriptions (selected)

Service name	Service Description	Data sources
MobiliCity	A smartphone application that aims to serve mobility-impaired users “by crowd-sourcing data…to improve London’s transport network.” Volunteer users’ movements around the city are acquired to “be able to infer where the ‘sticking-points’ are”. The insights are used “to report to authorities responsible for areas of difficulty, including TfL, Network Rail and London Boroughs with suggested improvements.”	[18], [53]
NY Housing Fire Control	After 2011, NYC built a catastrophic risk model to prevent fires in illegally converted housing spaces. “The City’s single largest source of intelligence on illegal conversions is New Yorkers who phone in…with tips” (p. 188). “By conducting an analysis of historic outcomes…we were able to create a risk model that takes each inbound 311 illegal conversion complaint … and predicts whether or not the complaint is most likely to be founded, meaning there are severely bad living conditions.” (p. 189) “That’s where we should be sending inspectors immediately.” (p. 189)	[51]

Eldesmarque InstaSport	A service that "[captures] real-time data from social media feeds to detect events that were happening around the city." Since Twitter is the main platform for InstaSport, they then "encapsulate the tweet and try to classify it into 4 different categories…we process the patterns searching for specific keywords in order to categorize the tweets. We developed our own solution for NLP". The identified events are "then added in another app or service, for example to warn about heavy traffic around a stadium or display available parking spots because people can't easily find it when they want to watch a match." (CEO, personal communication, March 2017)	[18]
Colour-in City	Colour-in City co-designed a service "that helps reduce stress and anxiety and improves wellbeing". The team works "in partnership with Lambeth Early Action partnership (LEAP), Colour-in City used chat bots to gather stories and perceptions of what it feels like to be a parent living in social housing in Lambeth." The service uses "AI chat bot to collect [subjective wellbeing] data", and use these insights where it "could help the council better support parents in social housing in the area." This is achieved through "chat bot integration with a private messaging service."	[18], [46], [54]
AirPublic	"The AirPublic network of 100–1000s of mobile sensors will provide… the necessary detailed information for micro-urban interventions, traffic routing systems and other potential applications of these datasets." Measurements using these sensors are more accurate by modelling the data while accounting for features related to "hyperlocal urban microclimates, such as the…street canyon effect."	[55]
Waze	"… community-based traffic and navigation app" where "users just drive with the app open on their phone to passively contribute traffic and other road data…or take a more active role by sharing road reports on accidents, police traps, or any other hazards along the way" "Waze is also home to an active community of online map editors who ensure that the data in their areas is as up-to-date as possible."	[56]
Public Like Displays (PLD)	PLD aims to "better communicate data on city challenges – such as waste management, energy consumption and changing demographics – to citizens." It takes the form of a product-service bundle, or a "toolbox that empowers citizens to publicly visualise [urban] data". The 3D printed toolboxes typically consisted of "… 1 interactive display that allows voting on a particular question, and 4 non-interactive displays that show related data, viewpoints or suggestions."	[18], [57]
Data-Waze	"Data-Waze came out of the desire to tell the stories… that utilize…data that our users provide." "…we can help people understand the reasons behind [traffic] and how it affects driving on a global scale." The narrative visualizations – in blog form on their main website – are developed around analysis of major traffic events, severe weather and studying its effect on driving or providing in-depth insights on specific cities.	[31]
Calimoto	Calimoto is a motorcycle app that uses a "number of different sensors in a Smartphone…[to] record your riding style... your angle in a particular curve or your average speed on a particular stretch. This enables every motorbike rider to have a thorough analysis of their own driving style in order to get more out of their bike and themselves." The application also uses the individual riding styles… "Based on our special algorithm… perfect route is calculated depending on [your] travel profile."	[58]
Google Traffic	"The traffic/speed database is built from multiple sources, including users. As the application is used on some phones, it reports its GPS information back to Google servers." The real-time traffic data and analytics technology comes from ZipDash, a startup Google acquired in 2005, and the users can see it as a "feature on Google Maps".	[59], [60]